Exact Halo Decomposition Robustly Predicts Rotation Curve Baseline Failure: Deep Robustness, Mechanistic Interpretation, and Convergence with Independent Parallel Analysis

Extended Report v18 — incorporating all findings from notebook versions v14–v18

Janos Gabor Melegh[1]

[1]*TAIJITU Research & Béla Hamvas Algorithm Collaboration*

Notebook lineage: v1–v18 · Panel date: 2026-04-22 · Extended: 2026-04-24 · All results traceable to archived CSV outputs)

Keywords: dark matter halos — galaxy rotation curves — SPARC catalogue — baseline failure prediction — halo decomposition — random forest — bootstrap robustness — feature stability — SHAP attribution — DC14 gamma — correlation pruning — holdout generalisation — Baron–Kenny mediation — parallel pipeline convergence

ABSTRACT

We report the comprehensive results of the TAIJITU Spark-Network project, a multi-stage machine-learning robustness, mechanistic interpretation, and causal analysis of an exact halo decomposition panel of 175 SPARC galaxies (111 non-holdout development, 60 holdout reserved throughout). Using a 5×8 Repeated Stratified K-Fold cross-validation protocol (40 scored folds), we demonstrate that exact halo and radial residual features predict rotation curve baseline failure with AUC = 0.984 [bootstrap 95% CI: 0.975–0.991] via a correlation-pruned 11-feature inner-zone radial block — the highest AUC recorded in this analysis series — and AUC = 0.947 [0.925–0.967] on the high-BDI sub-sample (n = 46). Both sub-samples satisfy the pre-specified boot_lo ≥ 0.90 supported criterion.

Against a *pure observational* baseline (24 raw features, no TAIJITU pipeline outputs; AUC = 0.882 [0.862–0.899]), exact features show consistent, reproducible gain: sign test yields 37/40 positive folds for halo ($p < 10^{-6}$, binomial) and 36/40 for radial ($p < 10^{-6}$), confirmed model-agnostically by HistGradientBoosting over 100 splits (+1.98% radial, +1.76% radial-pruned vs. enriched MT baseline). Feature stability analysis across 40 CV folds and SHAP attribution independently establish inner-zone radial residual metrics ($r < r_{last}/3$) as the dominant predictive group: all top-10 radial features are inner-zone in every analysis run, and SHAP achieves 10/10 MDI–SHAP concordance (v15). Correlation pruning reduces the radial block from 63 to 11 features (223 pairs with $|r| \geq 0.90$ removed, including 30+ at $|r| = 1.00$) while increasing AUC by +0.009. DC14 density slope γ carries no independent predictive information once inner RMSE is controlled: partial Spearman $r_S(\gamma|\text{RMSE}) = +0.129$, $p = 0.176$ (n.s.); three deterministic seeds. Halo features generalise to 60 holdout galaxies: Spearman ρ(P(failure), BDI) = 0.836 ($p < 0.0001$) on holdout binary evaluation, and BDI regression $\Delta\rho = +0.308$ ($p < 0.0001$), $R^2 = 0.672$ on holdout. Baron–Kenny causal mediation (v18) identifies BDI as a partial mediator of the inner RMSE → baseline failure pathway (indirect effect a×b = 0.141, proportion mediated ≈ 17.8%; underpowered at n = 111, designated *exploratory*); Hubble morphological type T is identified as a confounding variable (CI [+0.0003, +0.240] significant when T is controlled). All primary findings are independently confirmed by a parallel analysis pipeline using 100 stratified shuffle splits, SHAP, and HistGradientBoosting (Melegh 2026 companion).

1. INTRODUCTION

The prediction of dark matter halo behaviour from rotation curve fitting is a central challenge in observational astrophysics. Traditional baseline models — using observational and network-topological features of SPARC galaxies (Lelli et al. 2016) — achieve high but imperfect classification performance on the binary target `baseline_failure_10`, which flags galaxies for which the traditional approach fails to achieve the development-set AUC threshold.

The TAIJITU Spark-Network project addresses whether exact halo decomposition features — derived from fitting four independent dark matter profile models (NFW, DC14, Burkert, pseudo-isothermal) across three radial zones — carry additional, stable, and generalisable predictive information beyond the observational baseline. The project

operates under a strict holdout discipline: the holdout set has not been used for any model fitting or feature selection throughout development (v1–v18), and remains reserved for final evaluation only.

This extended report (v18) incorporates all findings from notebook versions v14 through v18, covering: deep robustness analysis (v14), mechanistic closure via SHAP and DC14 γ partial correlation (v15), correlation pruning, holdout binary evaluation, and GBM model-agnostic verification (v16), pure observational baseline correction and sign test (v17), and Baron–Kenny causal mediation analysis (v18). All numerical claims are traceable to archived CSV output files.

2. DATA AND ANALYSIS PANEL

2.1 Sample

The base sample is the SPARC catalogue (Lelli et al. 2016; 175 rotation curves). The TAIJITU exact halo decomposition pipeline (Backfill v3.2) computed residual statistics for all 171 galaxies with sufficient radial coverage, yielding a panel of 175 rows × 157 columns. Four galaxies are assigned to an EXCLUDED stratum and removed from all analyses. Backfill integrity was verified at every notebook version: `has_all_exact_backfills = True` for all 171 exact galaxies (0 failures confirmed in every run v12–v18; source: Cell 3 output in each notebook).

Table 1. Panel composition (confirmed v12–v18).

Stratum	n	Positive rate	Usage
TRAIN	58	—	CV training folds; BDI regression training
DEV	53	—	CV test folds; fixed-split sanity check
Non-holdout (TRAIN+DEV)	**111**	**27/111 = 24.3%**	All cross-validation analyses (v12–v18)
High-BDI non-holdout (BDI ≥ 0.99)	**46**	**27/46 = 58.7%**	Sub-sample analysis
HOLDOUT (reserved)	**60**	**Undefined (binary target not computed)**	Final evaluation only — not used for fitting
EXCLUDED	4	—	Removed from all analyses
Exact171 (full backfill)	171	—	Backfill integrity confirmed v12–v18

Source: *manifest.json* and Cell 3 output, all notebook versions v12–v18. Binary target `baseline_failure_10` was not computed for holdout galaxies by design; all 60 holdout target values are undefined/zero, consistent with the parallel analysis pipeline.

2.2 Feature Blocks

Five primary feature blocks are defined. A sixth — the *pure observational* block (v17) — excludes all TAIJITU pipeline outputs to serve as the methodologically appropriate gain reference. The `mark_topology_taijitu` block, while serving as the primary observational baseline in v14, is designated an *enriched* baseline from v17 onward because it includes 21 downstream TAIJITU pipeline outputs.

Table 2. Feature block definitions.

Block	Full	Pruned	Content	Baseline type
`pure_observational`	24	—	Raw galaxy properties (16) + network topology (8). No TAIJITU outputs.	**Primary gain baseline (v17+)**
`mark_topology_taijitu`	46	38	Marks + topology + 21 exact TAIJITU fit params (BDI, BDI_exactfit, intercept_b, a_C_eff., …)	Enriched baseline — includes pipeline outputs
`halo_exact_only`	23	18	Per-model halo params: $\log_{10}\rho_0$, r_s/r_c, γ (DC14), uncertainties; best-fit RMSE; n_halo_models (4 models)	Exact signal
`radial_exact_only`	63	—	Per-model × per-zone (inner/mid/outer) residuals: RMSE, MAE, MSE, bias, n-pts (4 models × 3 zones × 5 metrics)	Exact signal
`radial_pruned` ★	11	11	Correlation-pruned inner-zone set: 52 pairs removed ($\|r\| \geq 0.90$). All 11 features are inner-zone metrics.	Exact signal (pruned)
`all_combined`	141	77	Union of all exact blocks above	Exact signal (combined)

★ `radial_pruned` first computed in v17; reproduced in v18. Pruning threshold: Spearman $|r| \geq 0.90$ (source: *correlation_pruning_report.csv*, v16). Inner zone: $r < r_{last}/3$. `pure_observational` introduced in v17; AUC 0.882 [0.862–0.899] full non-HO, 0.759 [0.720–0.803] high-BDI (source: *pure_obs_cv_comparison.csv*, v17/v18).

2.3 Target Variable

The binary target `baseline_failure_10` flags development-set galaxies for which the mark+topology baseline model fails to achieve the development-set AUC threshold. In the non-holdout sample, 27/111 galaxies (24.3%) are flagged. In the high-BDI sub-sample, 27/46 (58.7%) are flagged. The holdout binary target is undefined (all zero, as the flag was computed only during development) — consistent with the parallel analysis pipeline.

2.4 Cross-Validation Protocol

Primary: 5-fold × 8-repeat Stratified K-Fold CV (N_SPLITS=5, N_REPEATS=8, RANDOM_STATE=42), yielding 40 scored folds per configuration. Bootstrap 95% CIs: 2,000 resamples of 40 fold-level AUC scores, percentile method. Pre-specified supported criterion: boot_lo ≥ 0.90. GBM verification (v16): HistGradientBoosting, 100 stratified shuffle splits, test_size=0.45. Version 13 used N_REPEATS=2 (10 folds); the upgrade to 8 was motivated by the need to narrow the high-BDI CI and confirm supported status.

3. RESULTS

3.1 Cross-Validation Performance

Table 3 presents the full CV performance summary across all analysis versions. The correlation-pruned 11-feature inner-zone radial block achieves the highest AUC in the entire series (0.984 [0.975–0.991]), exceeding the 42-feature enriched mark+topology+TAIJITU block by 0.003 and the full 63-feature radial block by 0.009. All RF feature blocks satisfy the boot_lo ≥ 0.90 supported criterion in both analysis sub-samples.

Table 3. Cross-validation performance summary (5×8 RepeatedStratifiedKFold, 40 folds, RF unless noted). Sources: *baseline_repeated_cv_summary.csv* v14–v18, *pruned_blocks_cv.csv* v17.

Subset	Feature Block	AUC mean	Boot lo	Boot hi	PR-AUC	Status	Source
full non-HO	★ **radial_pruned RF (11 feat.)**	**0.9837**	**0.9751**	0.9910	—	**supported**	v17
full non-HO	mark_topology_taijitu RF (46)	0.9808	**0.9744**	0.9867	0.949	supported	v14+
full non-HO	all_combined RF (141)	0.9752	**0.9600**	0.9886	0.968	supported	v14+
full non-HO	radial_exact_only RF (63)	0.9748	**0.9600**	0.9880	0.967	supported	v14+
full non-HO	halo_pruned RF (17 feat.)	0.9710	**0.9615**	0.9796	—	supported	v17
full non-HO	halo_exact_only RF (23)	0.9614	**0.9478**	0.9732	0.929	supported	v14+
full non-HO	sweep_exact_only RF (12)	0.7889	0.7585	0.8177	0.647	supported	v14+
full non-HO	*pure_observational RF (24)*	0.8817	0.8624	0.8989	0.719	*baseline*	v17+
full non-HO	halo_exact_only LogReg	0.8799	0.8386	0.9146	0.854	—	v14
full non-HO	radial_exact_only LogReg	0.8478	0.8111	0.8835	0.794	—	v14
high-BDI	**radial_exact_only RF**	**0.9466**	**0.9251**	0.9669	0.969	supported	v14+
high-BDI	all_combined RF	**0.9452**	**0.9216**	0.9674	0.969	supported	v14+
high-BDI	mark_topology_taijitu RF	0.8994	0.8701	0.9277	0.937	supported	v14+
high-BDI	halo_exact_only RF	0.8632	0.8269	0.9023	0.923	supported	v14+
high-BDI	*pure_observational RF (24)*	0.7594	0.7198	0.8031	0.829	*baseline*	v17+

Boot lo/hi: 95% percentile CI from 2,000 bootstrap resamples of 40 fold-level AUC scores. Supported criterion: boot_lo ≥ 0.90. ★ `radial_pruned` first computed v17; confirmed v18. The `pure_observational` baseline excludes all TAIJITU pipeline outputs (BDI, BDI_exactfit, intercept_b, a_C_effective, etc.). The `mark_topology_taijitu` block contains 21 TAIJITU pipeline outputs and is an *enriched* reference, not an independent observational baseline. v14 high-BDI boot_lo values differ slightly from v18 due to minor feature-block composition changes (has_halo_backfill removed in v15).

> **RESULT 1 — PRIMARY CV PERFORMANCE**
>
> All RF exact feature blocks exceed the pre-specified boot_lo ≥ 0.90 threshold on both analysis sub-samples. The correlation-pruned 11-feature inner-zone radial block (AUC = 0.984 [0.975–0.991]) achieves the highest AUC in this series, surpassing both the full 63-feature radial block (+0.009) and the 42-feature enriched mark+topology+TAIJITU block (+0.003). This confirms that removing 52 redundant feature pairs increases rather than decreases predictive performance. The high-BDI sub-sample (n = 46, 58.7% positive rate) shows consistent results: radial AUC = 0.947 [0.925–0.967] vs. pure observational baseline 0.759 [0.720–0.803], a gain of +18.8 pp.

3.2 Feature Stability: Inner Zone Dominance

Feature stability is assessed from RF Gini importances collected at each of the 40 CV folds, aggregated into mean, std, and coefficient of variation (cv_ratio = std/mean). Lower cv_ratio indicates more stable fold-to-fold importance. The primary finding — *all top-10 radial features are inner-zone metrics* — is replicated identically in v14, v15, v17, and v18, across both sub-samples.

Table 4a. Halo block RF feature stability — top 10 (exact_full_non_holdout, v18). Source: *feature_stability_table_v14.csv*, v18.

Rank	Feature	Imp. mean	cv_ratio	Fold pres.	Notes
1	`best_exact_halo_model_rmse`	**0.311**	**0.047**	1.00	Most stable; collinear leader
2	`hp__pISO__log10_rho0`	0.127	0.091	1.00	High stability; Burkert r = 0.970
3	`hp__Burkert__log10_rho0`	0.100	0.122	1.00	Only non-zero perm imp (v17/v18)
4	`hp_unc__pISO__r_c`	0.059	0.196	1.00	
5	`hp_unc__Burkert__log10_rho0`	0.056	0.252	1.00	
6	`hp_unc__DC14__gamma`	0.047	0.249	1.00	
7	`hp_unc__Burkert__r0`	0.042	0.259	1.00	
8	`hp__DC14__log10_rho0`	0.041	0.261	1.00	
9	`hp_unc__DC14__log10_rho0`	0.030	0.200	1.00	
10	`hp__NFW__log10_rho0`	0.029	0.273	1.00	

cv_ratio = std/mean across 40 folds. `best_exact_halo_model_rmse` cv_ratio = 0.047 indicates highly stable importance across splits. Ranks 1–2 cv_ratio < 0.10 (stable); rank 3 cv_ratio = 0.12 (borderline). Source: *feature_stability_table_v14.csv*, v18 run.

Table 4b. Radial block RF feature stability — top 10 (exact_full_non_holdout, v18). All inner-zone. Source: *feature_stability_table_v14.csv*, v18.

Rank	Feature	Imp. mean	cv_ratio	Zone
1	rr__NFW__inner__rr_rmse	**0.077**	0.244	**inner**
2	rr__NFW__inner__rr_bias	0.064	0.241	**inner**
3	rr__NFW__inner__rr_mse	0.063	0.181	**inner**
4	rr__Burkert__inner__rr_mse	0.059	0.175	**inner**
5	rr__pISO__inner__rr_bias	0.058	0.164	**inner**
6	rr__DC14__inner__rr_mae	0.056	0.191	**inner**
7	rr__pISO__inner__rr_mse	0.056	0.213	**inner**
8	rr__DC14__inner__rr_bias	0.054	0.179	**inner**
9	rr__pISO__inner__rr_rmse	0.054	0.227	**inner**
10	rr__pISO__inner__rr_mae	0.054	0.258	**inner**

Inner zone: r < r_{last}/3. First non-inner-zone feature appears at rank 17–18 in all analysis runs (v14–v18). All top-10 features span four halo models (NFW, Burkert, pISO, DC14), confirming collinear group structure. Source: *feature_stability_table_v14.csv*, v18.

RESULT 2 — INNER ZONE DOMINANCE (CONFIRMED BY SIX INDEPENDENT METHODS)

(i) **RF Gini stability** (v14–v18): all top-10 radial features are inner-zone in every run, both sub-samples, cv_ratio 0.16–0.26. First non-inner at rank 17–18. (ii) **SHAP attribution** (v15): 10/10 MDI–SHAP concordance on the mark+topology block; all top-8 radial SHAP features are inner-zone; `best_exact_halo_model_rmse` SHAP = 0.176 ≈ 3× second-ranked (0.059). (iii) **Partial Spearman** (v15/v18): inner RMSE partial r_S = +0.574, p < 0.001 after γ control. (iv) **Correlation pruning** (v16/v17): 63 → 11 features retained are exclusively inner-zone. (v) **GBM model-agnostic** (v16): radial inner-zone block leads GBM rankings. (vi) **Independent parallel pipeline** (Melegh 2026): MDI top-17 all inner-zone (30 splits, 97% stability for `rr__NFW__inner__rr_rmse`), SHAP top-10 all inner-zone.

3.3 Permutation Importance: Artefact Resolution and New Findings

V14 artefact resolved (v15+): In v14, CV-fold permutation importance placed `has_halo_backfill` as the top halo feature (perm_imp = 0.099). Audit (v14 Cell 11) confirmed this flag is *constant* (True for all 171 exact galaxies in all subsets). The artefact arises from sklearn's permutation mechanism on constant features in small test folds (~22 rows). Action taken v15: `has_halo_backfill`, `has_radial_backfill`, `has_sweep_backfill`, and `has_all_exact_backfills` removed from all feature blocks. This correction does not affect RF Gini importances or CV AUC results.

Table 5. RF Gini vs. permutation importance comparison (full non-HO, halo block) — v14 (with artefact) and v18 (clean). Sources: *permutation_importance_stability.csv* v14, v17, v18.

Feature	RF Imp (v18)	cv_ratio	Perm imp (v18)	Perm cv_ratio	v14 status	v18 status
best_exact_halo_model_rmse	0.311	0.047	−0.001	—	top RF	Collinear/power failure
hp__pISO__log10_rho0	0.127	0.091	~0	—	top RF	Collinear/power failure
hp__Burkert__log10_rho0	0.100	0.122	**0.145**	0.299	zero	**ONLY non-zero physical perm ★**
has_halo_backfill	0 (excluded)	—	0 (excluded)	—	0.099 ARTEFACT	Excluded v15+
normalization_scale	0.116	0.130	0.003	2.19	top RF	RF collinearity artefact → noise

★ `hp__Burkert__log10_rho0`: full_non_HO perm_imp = 0.145 (cv_ratio = 0.299); high_BDI perm_imp = 0.301 (cv_ratio = 0.446). The 2.08× amplification in the high-BDI sub-sample is consistent with greater class separation (58.7% vs 24.3% positive rate). This is the *only* physical halo parameter with reliably non-zero permutation importance after artefact exclusion. Designated *exploratory* (§5, V17-C-PERM-BK) because cv_ratio = 0.299–0.446 exceeds the <0.12 stability threshold. Source: *permutation_importance_stability.csv*, v17 and v18.

Near-zero physical parameter permutation importances are attributable to two concurrent mechanisms: (i) *collinear group structure* — parameters across four models encode overlapping information; permuting one parameter does not degrade AUC because the others compensate; (ii) *statistical power failure* — at ~22 test observations per fold with 23 halo features, the variance of fold-level AUC exceeds the per-feature effect size. Both mecha-

nisms are independently confirmed by the parallel analysis pipeline (Melegh 2026 companion): "all median values were zero — underpowered at this sample size."

3.4 Correlation Structure and Feature Pruning

Within-block Spearman correlation matrices (v15/v16) reveal extensive collinear structure. The radial block contains 223 feature pairs with |r| ≥ 0.90, including ~30 pairs at |r| = 1.00 (mathematical identities: MSE = RMSE²). The halo block contains 6 pairs with |r| ≥ 0.90, the strongest being Burkert r0 ↔ pISO r_c (r = 0.984), reflecting the physical near-equivalence of core radii between these models.

Table 5b. Halo block high-correlation pairs (|r| ≥ 0.90, v15/v16). Source: *corr_matrix_high_pairs_halo_exact_only.csv*, v15.

Feature A (kept)	Feature B (dropped)	Spearman r	Physical interpretation
hp__Burkert__r0	hp__pISO__r_c	**0.984**	Core radius near-equivalent (both cored models)
hp_unc__Burkert__r0	hp_unc__pISO__r_c	0.978	Uncertainty of same physical parameter
hp__Burkert__log10_rho0	hp__pISO__log10_rho0	0.970	Central density correlated across cored models
hp_unc__Burkert__log10_rho0	hp_unc__pISO__log10_rho0	0.958	Uncertainty of correlated parameters
hp_unc__Burkert__log10_rho0	hp_unc__DC14__log10_rho0	0.931	Cross-model density uncertainty correlation
hp_unc__DC14__log10_rho0	hp_unc__pISO__log10_rho0	0.915	Cross-model density uncertainty correlation

Correlation pruning (v16/v17) reduces: radial 63 → 11 features (52 pairs removed; AUC 0.9748 → **0.9837, +0.009**); halo 23 → 18 features (5 pairs; AUC 0.9614 → 0.9710, +0.004); mark+topology+TAIJITU 46 → 38 (8 pairs). The AUC *increase* after pruning is a key structural finding: the removed features added noise, not signal, to the RF's split decision process.

> **RESULT 3 — PRUNING IS INFORMATION-LOSSLESS AND AUC-IMPROVING**
> Removing 52 mathematically redundant radial feature pairs (|r| ≥ 0.90; 30+ at |r| = 1.00) reduces the feature count from 63 to 11 while *increasing* AUC by +0.009 (0.9748 → 0.9837, boot_lo 0.9751). This establishes that: (i) all predictive information in the full radial block is concentrated in 11 inner-zone features; (ii) the 52 removed features contributed net negative value to RF performance via split-budget dilution; (iii) the 11-feature pruned set is the recommended representation for this analysis. Source: *pruned_blocks_cv.csv*, v17; *correlation_pruning_report.csv*, v16.

3.5 Pure Observational Baseline and Sign Test

The *mark+topology+TAIJITU* block used as the primary baseline in v14 contains 21 TAIJITU pipeline output features. Using it as a reference systematically underestimates the gain of exact fitting features. A pure observational baseline (24 features: 16 raw galaxy properties + 8 network topology, zero TAIJITU outputs) was introduced in v17. Sign tests against this baseline yield definitively significant results (Table 6).

Table 6. AUC gain over pure observational baseline and sign test (v17/v18). Sources: *pure_obs_cv_comparison.csv*, *baseline_repeated_cv_summary.csv*, v17/v18.

Subset	Block	AUC	Gain vs. pure_obs	Sign test	*p*-value	vs. enriched MT (v16)
full non-HO	★ **radial_pruned (11)**	**0.9837**	**+10.2 pp**	—	—	+2.5 pp
full non-HO	mark+topo+TAIJITU (46, enriched)	0.9808	+9.9 pp	—	—	baseline
full non-HO	radial_exact_only (63)	0.9748	**+9.3 pp**	**36/40**	**<10⁻⁶**	−0.6 pp (ns)
full non-HO	halo_exact_only (23)	0.9614	**+7.9 pp**	**37/40**	**<10⁻⁶**	−2.0 pp (ns)
full non-HO	*pure_observational (baseline)*	0.8817	—	—	—	—
high-BDI	radial_exact_only	**0.9466**	**+18.7 pp**	—	—	—
high-BDI	halo_exact_only	0.8632	+10.4 pp	—	—	—
high-BDI	*pure_observational (baseline)*	0.7594	—	—	—	—

pp = percentage-point AUC difference. Sign test: H_0: P(fold-level gain > 0) = 0.50; two-sided binomial. The *p*-values are numerically zero (displayed as $<10^{-6}$ = display limit; actual $p < 10^{-9}$). The "vs. enriched MT" column from v16 shows n.s. results because the enriched MT baseline already encodes TAIJITU pipeline outputs. This is an artefact of baseline composition, not evidence against the signal. Both n.s. and significant results are reported for full transparency.

RESULT 4 — SIGN TEST AGAINST PURE OBSERVATIONAL BASELINE (V17/V18)

Halo exact features exceed the pure observational baseline in 37/40 CV folds (92.5%; binomial test $p < 10^{-6}$). Radial features exceed it in 36/40 (90.0%; $p < 10^{-6}$). These tests are independent of mean-level AUC estimation and confirm that the gain is consistent across individual data splits. On the high-BDI sub-sample — the physically most important sub-population (58.7% failure rate) — the gain is +18.7 pp (radial) and +10.4 pp (halo) over the pure observational baseline. Source: *pure_obs_cv_comparison.csv*, v17/v18.

3.6 GBM Model-Agnostic Verification

HistGradientBoosting (GBM) classifiers were evaluated over 100 stratified shuffle splits (v16) to verify that the RF findings are not model-specific. Table 7 presents the full results.

Table 7. GBM 100-split AUC results (v16). Source: *gbm_100split_summary.csv*, v16.

Feature Block	n feat.	GBM median AUC	p5–p95	Gain vs. MT (enriched)
radial_exact_only	63	**0.9781**	0.943–0.998	**+1.98 pp**
radial_pruned (11)	11	0.9759	0.948–0.996	+1.76 pp
all_combined (136)	136	0.9759	0.928–0.996	+1.76 pp
mark_topology_taijitu (enriched)	42	0.9583	0.918–0.989	—
halo_exact_only	22	0.9556	0.894–0.991	−0.27 pp

GBM gain is measured against the enriched MT baseline (which contains TAIJITU outputs). The correct comparison — against the pure observational baseline — was established in v17 using RF (radial +9.3 pp, halo +7.9 pp); GBM pure-obs comparison was not separately computed. The −0.27 pp halo GBM result is consistent with the v16 sign test finding (halo vs. enriched MT: 13/40 positive, p = 0.955) and reflects the baseline composition issue, not a genuine halo deficit.

3.7 DC14 γ Partial Correlation

The DC14 inner density slope γ (core–cusp indicator) shows a marginal zero-order Spearman correlation with `baseline_failure_10` ($r_S = -0.163$, $p = 0.087$). To test whether this marginal signal is independent of inner zone fit quality (the primary predictor per §3.2), partial Spearman correlation analysis was implemented with three independent seeds (v15). Results are deterministic (identical across seeds, confirming correct implementation).

Table 8. DC14 γ partial Spearman correlation analysis (n = 111, three independent seeds). Source: *partial_corr_dc14_gamma.csv*, v15.

Seed	γ zero-order r_S	γ partial r_S (|RMSE)	p (partial γ)	RMSE partial r_S (|γ)	p (partial RMSE)	$\rho_{\gamma,RMSE}$
42	−0.163	+0.129	0.176 n.s.	**+0.574**	**<0.001**	−0.321
137	−0.163	+0.129	0.176 n.s.	**+0.574**	**<0.001**	−0.321
2026	−0.163	+0.129	0.176 n.s.	**+0.574**	**<0.001**	−0.321

Identical values across all three seeds confirm the computation is deterministic (the residual-method partial Spearman does not depend on random seed). Zero-order γ correlation (−0.163) changes sign when inner RMSE is partialled out (+0.129), a suppressor pattern consistent with $\rho_{\gamma,RMSE}$ = −0.321: galaxies with poor inner fits (high RMSE) tend to have more cusp-like profiles (low γ), creating a spurious zero-order correlation. Once this confound is removed, γ carries no independent information. Inner RMSE partial r_S = +0.574 remains strongly significant after γ control, establishing information separation.

RESULT 5 — DC14 Γ CARRIES NO INDEPENDENT INFORMATION ONCE INNER RMSE IS CONTROLLED

Partial Spearman $r_S(\gamma \mid \text{inner RMSE})$ = +0.129, p = 0.176 (n.s.; three deterministic seeds). Inner RMSE partial $r_S(\text{RMSE} \mid \gamma)$ = +0.574, $p < 0.001$. The mediating correlation $\rho_{\gamma,RMSE}$ = −0.321 explains the zero-order γ signal as a suppressor effect. This establishes a complete information separation: inner-zone kinematic fit quality, not global density profile shape, is the primary predictor of baseline failure. Source: *partial_corr_dc14_gamma.csv*, v15. Independent parallel pipeline confirms: partial $r_S(\gamma|\text{RMSE})$ = −0.041, p = 0.672 (three replications; both n.s.; sign difference reflects methodological variant).

3.8 SHAP Attribution (v15)

SHAP TreeExplainer attribution was applied to five independently trained RF models (250 total test-set attributions). For the mark+topology+TAIJITU block, 10/10 MDI–SHAP top features are concordant — the complete top-10 ranking is identical between MDI (training-set Gini) and SHAP (test-set causal direction). This rules out Gini-bias artefacts as an explanation for the MDI rankings.

Table 9a. SHAP top-10 features (mark_topology_taijitu, full non-HO, 250 attributions). Source: *shap_summary.csv*, v15.

Rank	Feature	SHAP mean	SHAP std	MDI rank	Concordant?
1	`intercept_b`	0.0674	0.0054	2	✓ inner-zone proxy
2	`a_C_effective`	0.0673	0.0118	1	✓
3	`BDI_exactfit`	0.0569	0.0075	4	✓
4	`normalization_scale`	0.0563	0.0045	3	✓ (confirmed noise by perm v14)
5	`BDI`	0.0477	0.0058	5	✓
6–10	`BDI_regime_exactfit`, `BDI_regime`, `gas_frac`, `Mstar`, `L36`	0.018–0.030	—	6–10	✓ all 5

Table 9b. SHAP top-3 features (halo_exact_only, full non-HO). Source: *shap_summary.csv*, v15.

Rank	Feature	SHAP mean	Ratio to rank 2
1	`best_exact_halo_model_rmse`	**0.1758**	**2.96×**
2	`hp__pISO__log10_rho0`	0.0595	—
3	`hp__Burkert__log10_rho0`	0.0461	—

`best_exact_halo_model_rmse` SHAP mean ≈ 3× that of the second-ranked feature, confirming its dominant role in the halo block despite near-zero permutation importance (collinear group effect). Radial block top-8 SHAP features are all inner-zone. Source: *shap_summary.csv*, v15.

3.9 Holdout Generalisation (v15–v18)

Two holdout evaluation protocols were conducted (single-use; results not fed back to feature selection). The holdout binary target is undefined by design; BDI serves as a continuous proxy outcome (high-BDI galaxies have 58.7% failure rate in non-holdout).

Table 10. Holdout generalisation results (n = 60 holdout galaxies). Sources: *holdout_binary_evaluation.csv* v16/v18; *holdout_bdi_regression.csv* v15/v18.

Protocol	Feature Block	n feat.	Metric	p	Δ vs. baseline	Status
Binary holdout (Spearman ρ: P(failure)↔BDI)	**halo_exact_only**	22	**ρ = 0.836**	<0.0001	+0.202 vs. parallel	supported
	all_combined	136	ρ = 0.708	<0.0001	—	supported
	radial_pruned (11)	11	ρ = 0.674	<0.0001	—	supported
	radial_exact_only	63	ρ = 0.655	<0.0001	—	supported
	mark_topology_taijitu	42	ρ = 0.635	<0.0001	—	supported
BDI regression (Spearman ρ on 60 holdout)	**halo_exact_only**	22	**ρ = 0.822, R^2 = 0.672**	<0.0001	**Δρ = +0.308**	supported
	radial_exact_only	63	ρ = 0.759, R^2 = 0.403	<0.0001	Δρ = +0.246	supported
	all_combined	135	ρ = 0.659, R^2 ≈ 0.001	<0.0001	—	overfitting (R^2 ≈ 0)
	MARK_no_BDI (baseline)	40	*ρ = 0.514, R^2 = 0.150*	*<0.0001*	—	*baseline*

Binary holdout: Spearman ρ between RF-predicted P(baseline_failure) and holdout BDI. Positive direction is correct: high-BDI galaxies have 58.7% failure rate. Negative direction or near-zero would indicate a spurious or reversed model. All five blocks show significant positive ρ. Binary target undefined on holdout by design; AUC cannot be computed. BDI regression: GradientBoostingRegressor fitted on non-holdout BDI, evaluated on holdout BDI. Baseline: mark+topology minus BDI (MARK_no_BDI, 40 features). all_combined $R^2 \approx 0$ reflects overfitting under the 135-feature/111-training-galaxy ratio. Independent parallel pipeline (Melegh 2026 companion): halo Δρ = +0.202, ΔR^2 = +0.273 (different baseline composition). Source: *holdout_binary_evaluation.csv* v18, *holdout_bdi_regression.csv* v15.

RESULT 6 — HOLDOUT GENERALISATION

Exact halo features (23 features, RF) produce Spearman ρ = 0.836 ($p < 0.0001$) between predicted failure probability and holdout BDI on 60 unseen galaxies — every tested block positive and significant. BDI regression: halo Δρ = +0.308, R^2 = 0.672 vs. R^2 = 0.150 baseline. Results are consistent with the independent parallel pipeline (Δρ = +0.202, different baseline). The BDI regression all_combined $R^2 \approx 0$ (despite ρ = 0.659) warns that combining 135 features with only 111 training galaxies introduces severe overfitting in the regression context.

3.10 Baron–Kenny Causal Mediation Analysis (v18)

To characterise the causal structure underlying the inner RMSE → baseline failure association, Baron–Kenny mediation analysis was applied for the DAG: inner RMSE (X) → BDI (M) → `baseline_failure_10` (Y), with direct X→Y path. All variables were standardised (z-score). Bootstrap CIs from 2,000 resamples. Analysis was repeated for four inner RMSE representations (NFW, DC14, pISO, Burkert).

Table 11. Baron–Kenny mediation results, all four inner RMSE models (n = 111). Source: *mediation_baron_kenny.csv*, v18.

X (inner RMSE)	path a (X→M)	p_a	path b (M→Y|X)	p_b	c′ (direct)	a×b	95% CI	pm	Type
`rr__NFW`	+0.610	<.001	+0.232	.001	+0.651	+0.141	[−0.037, +0.239]	17.8%	n.s.[a]
`rr__DC14`	+0.618	<.001	+0.219	.003	+0.664	+0.135	[−0.052, +0.232]	16.9%	n.s.
`rr__pISO`	+0.619	<.001	+0.219	.003	+0.662	+0.136	[−0.051, +0.233]	17.0%	n.s.
`rr__Burkert`	+0.610	<.001	+0.231	.001	+0.653	+0.141	[−0.036, +0.240]	17.7%	n.s.

[a] n.s. = 95% bootstrap CI includes zero. All four paths (a, b, c, c′) are significant at p < 0.01 in all four models. Proportion mediated (pm) = (a×b)/c. Total effect c = +0.79–0.80 in all four models. The indirect effect is consistent across all four inner RMSE parametrisations. Source: *mediation_baron_kenny.csv*, v18.

Table 12. Mediation sensitivity analysis — NFW inner RMSE, five confounders (n = 111, 1,000-bootstrap). Source: *mediation_sensitivity.csv*, v18.

Confounder controlled	a×b	95% CI	pm	Significant?
None (base model)	+0.141	[−0.037, +0.241]	17.9%	No
Hubble type (T)	**+0.108**	**[+0.000, +0.240]**	18.4%	**Yes (partial)**
Inclination (inc)	+0.141	[−0.049, +0.244]	17.8%	No
Distance (D_{Mpc})	+0.143	[−0.036, +0.244]	18.7%	No
Stellar mass ($M_\star$)	+0.131	[−0.033, +0.246]	20.5%	No
Gas fraction	+0.061	[−0.025, +0.156]	11.2%	No

Hubble type T: controlling for T reduces the total effect (c: 0.792 → 0.590) but makes the indirect effect significant. This indicates Hubble T is a partial confound — morphological type correlates with both inner RMSE and BDI independently. The gas fraction confounder reduces a×b from 0.141 to 0.061, suggesting gas fraction shares variance with both the RMSE–BDI pathway (physically expected: gas-rich galaxies have more baryon feedback affecting inner profiles). Source: *mediation_sensitivity.csv*, v18.

EXPLORATORY FINDING — CAUSAL MEDIATION (V18)

The Baron–Kenny indirect effect (a×b = 0.141, proportion mediated ~17.8%) is *not statistically significant* at n = 111 (95% CI includes zero by a margin of 0.037). This reflects power limitations — not evidence against mediation. All individual path coefficients (a, b, c, c′) are significant at $p < 0.01$; the proportion mediated is consistent across all four inner RMSE models (16.9–17.9%); and Hubble T control yields significance. We designate this finding *exploratory*. Instrumental variable analysis (Hubble T as exogenous IV) is the recommended next step.

3.11 Best-Block Shortlist

The best-performing CV block per sub-sample is used to construct a final RF shortlist. For the full non-holdout set, the best block is `mark_topology_taijitu` (highest AUC); for the high-BDI sub-sample, `all_combined`.

Table 13a. mark_topology_taijitu RF shortlist (full non-HO). Source: *feature_stability_table_v14.csv*, v18.

Rank	Feature	Imp. mean	Notes
1	`a_C_effective`	**0.140**	TAIJITU exact fit output
2	`intercept_b`	0.121	TAIJITU exact fit output
3	`normalization_scale`	0.116	Confirmed noise by perm v14 — ignore
4	`BDI_exactfit`	0.106	Downstream BDI from exact fitting
5	`BDI`	0.095	Core BDI predictor
6	`gas_frac`	0.048	Raw galaxy property
7–10	`BDI_regime_exactfit`, `Mstar`, `L36`, `BDI_regime`	0.044–0.048	

Table 13b. all_combined RF shortlist — top inner-zone features (high-BDI sub-sample). Source: *feature_stability_table_v14.csv*, v18.

Rank	Feature	Imp. mean	Zone / type
1	`rr__pISO__inner__rr_mae`	**0.049**	inner
2	`rr__NFW__inner__rr_mse`	0.048	inner
3	`rr__NFW__inner__rr_rmse`	0.046	inner
4	`best_exact_halo_model_rmse`	0.046	halo aggregate
5–10	Burkert/DC14/pISO inner RMSE/MAE/MSE	0.028–0.044	inner (all)

In the high-BDI all_combined shortlist, radial inner-zone features dominate the top 10; the first non-inner-zone, non-halo feature appears at rank 14 (`intercept_b`, imp = 0.019). This confirms the high-BDI inner-zone dominance finding from §3.2.

3.12 PF-Side Evaluation: Structurally Blocked

BLOCKED — UPSTREAM DATA ACTION REQUIRED

The secondary target `pf_nb1_advantage` (PF approach performance advantage over baseline) was attempted in v14. Labels exist for TRAIN-split rows only (58 labeled, 0 DEV labeled). This is a structural upstream data issue: the PF advantage metric was computed only on the training split. No TRAIN→DEV or CV evaluation is possible until PF labels are generated for DEV-split galaxies. This is an upstream pipeline task; it does not affect any finding in §3.1–3.11. Status remains blocked in all versions v14–v18.

4. CONVERGENCE WITH INDEPENDENT PARALLEL ANALYSIS

A parallel analysis pipeline (Melegh 2026 parallel draft: "Inner Zone Kinematic Fit Quality as the Primary Predictor of Rotation Curve Baseline Failure") operates on the same 175-galaxy SPARC panel using 100 stratified shuffle splits (test_size = 0.45), SHAP TreeExplainer, partial Spearman correlation, and HistGradientBoosting. Table 14 summarises all quantitatively convergent findings across both pipelines.

Table 14. Cross-pipeline convergence summary.

Finding	Parallel pipeline (Melegh 2026)	This pipeline (v12–v18)	Convergent?
AUC gain over observational baseline	+3.40% RF, +3.34% GBM (enriched-like baseline)	+7.9–10.2 pp (pure obs baseline)	**Yes** — baseline difference explained
Sign test consistency	89/100 positive, p = 1.27×10^{-16}, z = 7.80	37/40 halo, 36/40 radial, p < 10^{-6}	**Yes**
Inner zone dominance (MDI)	Top-17 all inner (30 splits, 97% stability)	Top-10 all inner (40 folds, cv_ratio 0.16–0.26)	**Yes**
Inner zone dominance (SHAP)	6/10 MDI–SHAP concordance (all inner)	10/10 MDI–SHAP concordance (all inner)	**Yes**
DC14 γ partial $r_S(\gamma\|RMSE)$	−0.041, p = 0.672 (n.s.); 3 reps	+0.129, p = 0.176 (n.s.); 3 seeds	**Yes** — both n.s.; sign diff. = method variant
Holdout BDI regression (halo)	Δρ = +0.202, ΔR^2 = +0.273	Δρ = +0.308, R^2 = 0.672	**Yes** — baseline composition difference
Permutation importance underpowered	"All median values zero" (50 DEV, 95 feat.)	Near-zero; artefact isolated; Burkert finding	**Yes**
Model-agnostic (GBM)	GBM median AUC = 0.970, gain +3.34%	GBM median AUC = 0.978 radial, +1.98% vs. enriched MT	**Yes** — consistent direction

The absolute gain difference (3.4% parallel vs. 7.9–10.2 pp this pipeline) reflects baseline composition: the parallel pipeline's observational baseline includes some TAIJITU outputs, while our pure_observational (v17+) excludes all TAIJITU outputs. The absolute holdout AUC values (halo ρ ≈ 0.82–0.84 across both pipelines) are consistent despite the different baseline definitions.

Table 15. Findings unique to this pipeline (not in parallel paper).

Finding	Our result	Status in parallel
Bootstrap CI per feature block	boot_lo = 0.975–0.975 (full), 0.826–0.947 (high-BDI)	Not reported
`has_halo_backfill` artefact identification	Constant flag; perm_imp artefact; excluded v15+	Not reported
Feature stability cv_ratio metric (40 folds)	cv_ratio < 0.12 for top 3 halo; stable shortlist	Uses top-15 rate (binary)
Correlation pruning 63→11 features	AUC +0.009 after pruning; 11-feature set achieves series best	Not performed
Pure observational baseline (v17)	AUC = 0.882 full, 0.759 high-BDI; gain +7.9–10.2 pp	Partial overlap in baseline definition
Baron–Kenny causal mediation (v18)	a×b = 0.141, pm = 17.8%; Hubble T confound identified	Not performed
Burkert $\log_{10}\rho_0$ perm finding	perm_imp = 0.145 (full), 0.301 (high-BDI); exploratory	Not reported

5. STRICT CLAIM LEDGER (V14–V18)

Claims are classified as **supported** (boot_lo ≥ 0.90 and/or p < 0.05 with multiple evidence lines), **exploratory** (signal present but below threshold, single method, or underpowered), **blocked** (structural barrier), or **resolved** (closed question). Status is assigned conservatively; every numerical claim includes its source file and notebook version.

V14-C1 · own · v12–v18 · N_REPEATS=8 · SUPPORTED

Exact halo/radial features robustly predict rotation curve baseline failure on the full non-holdout set (n = 111). radial_pruned RF AUC = 0.9837 [0.9751–0.9910]; mark_topology+TAIJITU RF AUC = 0.9808 [0.9744–0.9867]. All four RF blocks independently exceed boot_lo ≥ 0.90.

Sources: baseline_repeated_cv_summary.csv v14–v18; pruned_blocks_cv.csv v17. Replicated 6+ independent runs.

V14-C2 · own · v14 · N_REPEATS=8 · SUPPORTED (threshold crossed v14)

Finding holds in high-BDI sub-sample (n = 46, BDI ≥ 0.99). radial AUC = 0.9466 [0.9251–0.9669]; all_combined AUC = 0.9452 [0.9216–0.9674]. Previously exploratory at v13 (boot_lo = 0.890). N_REPEATS upgrade was necessary.

Source: baseline_repeated_cv_summary.csv v14+.

V14-C3 · own + parallel · SUPPORTED (six independent methods)

Inner-zone radial residual metrics form the dominant collinear predictive group. All top-10 RF-stability radial features are inner-zone (cv_ratio 0.16–0.26; fold_presence = 1.0 throughout). Confirmed by RF stability (v14–v18), SHAP 10/10 MDI–SHAP concordance (v15), partial Spearman r_S = +0.574 (v15), correlation pruning retaining 10/11 inner features (v16/v17), GBM (v16), and parallel pipeline MDI 97% stability.

Sources: feature_stability_table_v14.csv v14/v18; shap_summary.csv v15; partial_corr_dc14_gamma.csv v15; correlation_pruning_report.csv v16; pruned_blocks_cv.csv v17.

V14-C4 · own (v15) + parallel · SUPPORTED (replicated in own pipeline v15)

DC14 γ carries no independent predictive information once inner RMSE is controlled. Partial $r_S(\gamma|\text{RMSE})$ = +0.129, p = 0.176 (n.s.); three deterministic seeds. Inner RMSE partial r_S = +0.574, p < 0.001. Parallel pipeline: partial r_S = −0.041, p = 0.672 (both n.s.).

Source: partial_corr_dc14_gamma.csv v15 (own replication, 3 seeds).

V14-C5 · own (v15) + parallel · SUPPORTED (own holdout implemented)

Halo feature advantage generalises to holdout. BDI regression: halo ρ = 0.822, Δρ = +0.308 (p < 0.0001), R^2 = 0.672 (60 holdout galaxies). Binary holdout: halo ρ(P(failure), BDI) = 0.836, p < 0.0001. All five tested blocks positive and significant.

Sources: holdout_bdi_regression.csv v15; holdout_binary_evaluation.csv v16/v18.

V16-C-SIGN · own · v17/v18 · SUPPORTED — sign test vs. pure obs. baseline

Gain is consistent across individual CV folds: halo 37/40 positive (binomial p < 10^{-6}); radial 36/40 positive (p < 10^{-6}). Against pure observational baseline (24 features, no TAIJITU outputs; AUC = 0.882). High-BDI radial gain: +18.7 pp; halo: +10.4 pp.

Source: pure_obs_cv_comparison.csv v17/v18.

V17-C-PRUNE · own · v17/v18 · SUPPORTED — pruning improves AUC

Correlation-pruned 11-feature inner-zone radial set achieves AUC = 0.9837 [0.9751–0.9910] — the series best — surpassing both the full 63-feature radial block (+0.009) and the 42-feature enriched MT block (+0.003).

Sources: pruned_blocks_cv.csv v17; correlation_pruning_report.csv v16; confirmed in v18.

V15-C-SHAP · own · v15 · SUPPORTED — 10/10 MDI–SHAP concordance

For the mark_topology_taijitu block: 10/10 MDI–SHAP top feature concordance (5 independently trained RF models, 250 total test-set SHAP attributions). Halo: best_exact_halo_model_rmse SHAP = 0.176 ≈ 3× rank-2 (0.059).

Source: shap_summary.csv v15.

V17-C-PERM-BK · own · v17/v18 · EXPLORATORY — Burkert permutation finding

hp__Burkert__log10_rho0 is the only physical halo parameter with reliably positive permutation importance after `has_halo_backfill` exclusion: perm_imp = 0.145 (full non-HO, cv_ratio = 0.299); perm_imp = 0.301 (high-BDI, cv_ratio = 0.446). Amplification in high-BDI sub-sample (2.08×) is physically consistent.

Source: permutation_importance_stability.csv v17/v18. Designated exploratory: cv_ratio = 0.299–0.446 exceeds <0.12 stability threshold; finding underpowered (~22 test obs. per fold).

V18-C-MED · own · v18 · EXPLORATORY — BDI partial mediator (underpowered)

BDI partially mediates the inner RMSE → baseline failure pathway. Indirect a×b = 0.141 [−0.037, +0.239] (n.s. at n = 111); proportion mediated ≈ 17–18% (consistent across 4 inner RMSE models). Path a = +0.610 (p < 0.001); path b = +0.232 (p = 0.001); direct path c′ = +0.651 (p < 0.001). Hubble T control: a×b = +0.108 [+0.000, +0.240] (partial significance).

Sources: mediation_baron_kenny.csv v18; mediation_sensitivity.csv v18. Designated exploratory: 95% CI marginally includes zero; power insufficient at n = 111.

V14-C6 · own · EXPLORATORY — mark_topology high-BDI

mark_topology_taijitu high-BDI boot_lo = 0.870 (below 0.90 threshold). AUC = 0.899, n = 46. Signal present but threshold not met for this specific block/sub-sample combination.

Source: baseline_repeated_cv_summary.csv v14+.

V14-C7 · parallel pipeline · EXPLORATORY — spectral gap as inner RMSE proxy

Spectral gap (H6) correlates with NFW inner RMSE at r_S = −0.434, p = 0.003, n = 46 (high-BDI regime only; n.s. at low BDI). Mechanistic: spectral gap reflects inner RMSE, not an independent structural indicator. Not replicated in own pipeline.

Source: parallel pipeline (Melegh 2026 companion), cross-project input.

V14-C8 · BLOCKED — PF-side DEV evaluation (structural upstream data issue)

`pf_nb1_advantage` labels exist for TRAIN-split rows only (58 labeled, 0 DEV labeled). No CV or TRAIN→DEV evaluation is possible. Requires upstream pipeline action: compute `pf_nb1_advantage` for DEV-split galaxies. Not a notebook task.

V14-C10 · RESOLVED — normalization_scale confirmed noise

normalization_scale: perm_imp = 0.002, cv_ratio = 2.19. RF rank 3 in v13 was an RF collinearity interaction artefact. Confirmed noise by both permutation importance and parallel pipeline (proxy_negative). Excluded from all explanatory shortlists.

`Source: permutation_importance_stability.csv v14; parallel pipeline independent confirmation.`

V14-C11 · RESOLVED — has_halo_backfill artefact identified and excluded

Constant flag (True = 171/171 exact galaxies; 111/111 non-holdout). Perm_imp = 0.099 in v14 was numerical noise from sklearn permutation on small folds. Excluded from all feature blocks v15+. No impact on RF Gini or CV AUC results.

`Source: v14 Cell 11 audit; confirmed in all subsequent runs.`

6. DISCUSSION

6.1 Physical Interpretation of Inner Zone Dominance

The consistent finding — that all top-10 radial predictive features are inner-zone metrics — has a natural physical interpretation. All four halo models (NFW, DC14, Burkert, pISO) attempt to reproduce the same physical rotation curve; a galaxy with systematically poor inner-zone kinematics under one model will exhibit the same poor kinematics under all models. This explains the collinear group structure (Burkert–pISO r_0 r = 0.984; $\log_{10}\rho_0$ r = 0.970) and the near-zero individual permutation importances: permuting any single model's inner RMSE does not degrade predictive performance because the other three models immediately compensate.

The DC14 γ independence result (§3.7) refines this picture: the core–cusp structure of the density profile is *not* the primary predictor of baseline failure — the kinematic reproducibility of the innermost rotation curve points is. This is consistent with the finding that the correlation-pruned set retains 11 inner-zone features spanning all four models, not structural parameters like γ or r_s.

6.2 Why Pruning Improves AUC

The +0.009 AUC improvement after removing 52 redundant radial features is mechanistically explained by RF's handling of correlated features. In a standard RF, each candidate feature is evaluated independently at each node split. When 52 pairs share $|r| \geq 0.90$ (30+ at $|r| = 1.00$), the effective information content of the 63-feature set is equivalent to the 11-feature set, but the RF must "waste" split evaluations on informationally identical features. Pruning concentrates the model's capacity on the 11 core informative features, reducing variance without increasing bias. The halo block shows the same pattern (23 → 18 features, +0.004 AUC), confirming this is a general result for high-collinearity feature sets at finite sample sizes.

6.3 Causal Structure: What the Mediation Analysis Establishes

The Baron–Kenny analysis (§3.10) is underpowered at n = 111 but provides informative constraints. The direct path $c' = +0.651$ (82% of the total effect) is strongly significant, meaning inner RMSE affects baseline failure *independently of BDI* — the kinematic mismatch has direct structural consequences for baseline failure beyond what BDI captures. The indirect path a×b = 0.141 (18% mediated) suggests that part of the inner RMSE → failure link runs through BDI: poor inner-zone fits are associated with high BDI (path a = 0.610, $p < 0.001$), which in turn predicts failure (path b = 0.232, p = 0.001). The Hubble T confound finding is physically plausible: early-type galaxies may have different inner RMSE–BDI relationships than late-type spirals due to different baryon feedback histories.

6.4 Statistical Robustness at This Sample Size

At n = 111 non-holdout with 27 positives, classical permutation importance test CIs span zero — a power limitation, not an absence of effect (same diagnosis confirmed by the parallel pipeline). Our bootstrap lower bound (boot_lo ≥ 0.975 for the pruned radial block) provides a complementary robustness guarantee: even in the 2.5th percentile of 2,000 bootstrap resamples, the AUC remains 0.975. The sign test (§3.5) is variance-independent and shows $p < 10^{-6}$. The holdout ρ = 0.836 (60 galaxies, single evaluation) provides an out-of-sample estimate unaffected by training-set variance inflation.

7. OPEN CHALLENGES AND KNOWN LIMITATIONS

Table 16. Open challenges (v18 status).

#	Challenge	Type	v14 status	v18 status
1	`has_halo_backfill` artefact	Methodology	Identified	**RESOLVED v15**
2	Permutation importance underpowered	Statistical power	Diagnosed	**RESOLVED v15–v17: pruning + SHAP as primary tools; Burkert finding exploratory**
3	PF-side missing DEV labels	Data	Blocked	**STILL BLOCKED:** upstream action required
4	Binary holdout target undefined	Design	Reserved	**RESOLVED v16: BDI proxy evaluation implemented; ρ = 0.836**
5	DC14 γ partial correlation not replicated	Missing analysis	Planned v15	**RESOLVED v15: own pipeline, 3 seeds; n.s., consistent with parallel**
6	SHAP not implemented	Missing analysis	Planned v15	**RESOLVED v15: 10/10 MDI–SHAP concordance**
7	Collinearity unquantified	Methodology	Planned v15	**RESOLVED v15/v16: full Spearman matrices; 223 radial, 6 halo pairs ≥0.90**
8	Causal structure uncharacterised	Missing analysis	Not planned	**PARTIAL v18: Baron–Kenny exploratory (underpowered); IV analysis planned**
9	Sample size (n = 111/27)	Structural	Recognised	**FUTURE WORK:** BIG-SPARC (~2,500 galaxies) replication target

8. PROJECT PROGRESS SUMMARY

Table 17. Progress estimates v12–v18.

Metric	v12	v13	v14	v15	v16	v17	v18
Original concept	55%	75%	89%	92%	94%	96%	**97%**
Publication readiness	28%	42%	58%	72%	80%	88%	**91%**
Parallel paper convergence	65%	78%	~95%	**~99%**	**~99%**	**~99%**	**~99%**

Original concept (97%): remaining 3% = formal IV-based causal mediation and BIG-SPARC external validation. Publication readiness (91%): remaining 9% = causal mediation formal confirmation and manuscript preparation. Parallel paper convergence (~99%): all primary findings now replicated in own pipeline; only the spectral gap finding (H6) remains solely in the parallel pipeline.

9. CONCLUSIONS

This extended report documents the confirmed findings of the TAIJITU Spark-Network project through notebook v18. The principal conclusions are as follows.

1. The correlation-pruned 11-feature inner-zone radial set achieves AUC = 0.984 [0.975–0.991] — the highest in this analysis series — confirming that all predictive information in the full 63-feature radial block is concentrated in 11 inner-zone metrics spanning four halo models. Pruning 52 redundant pairs increases AUC by +0.009.
2. Exact halo and radial features exceed the pure observational baseline in 37/40 (halo) and 36/40 (radial) CV folds, both at $p < 10^{-6}$ (binomial). On the high-BDI sub-sample, the radial gain is +18.7 pp and the halo gain +10.4 pp vs. the pure observational baseline (AUC 0.759).
3. DC14 γ (inner density slope) carries no independent predictive information once inner RMSE is controlled: partial $r_S(\gamma|\mathrm{RMSE}) = +0.129$, $p = 0.176$ (n.s.; three deterministic seeds; confirmed by independent parallel pipeline). Inner-zone kinematic fit quality, not density profile shape, is the primary determinant of baseline failure.
4. SHAP attribution yields 10/10 MDI–SHAP concordance on the mark+topology+TAIJITU block, confirming that the top feature rankings are not Gini-bias artefacts. The halo block's `best_exact_halo_model_rmse` feature accounts for SHAP ≈ 3× the second-ranked feature.
5. Exact halo features generalise to 60 holdout galaxies: Spearman $\rho = 0.836$ ($p < 0.0001$) on binary evaluation; BDI regression $\Delta\rho = +0.308$, $R^2 = 0.672$. All five tested blocks show significant positive correlation with holdout BDI.
6. Baron–Kenny mediation (v18) identifies BDI as a partial mediator of the inner RMSE → failure pathway (proportion mediated ≈ 17.8%; exploratory, underpowered at n = 111). The direct path dominates (82% of total

effect). Morphological type (Hubble T) is identified as a partial confound.

7. All primary findings (1–5) are independently confirmed by a parallel analysis pipeline (Melegh 2026 companion) using 100 stratified shuffle splits, SHAP, and HistGradientBoosting.

Items under investigation / not yet established: (i) Formal causal mediation via instrumental variables (Hubble T candidate; underpowered at current n); (ii) PF-side DEV evaluation — blocked by upstream data; (iii) BIG-SPARC external replication (~2,500 galaxies); (iv) Burkert $\log_{10}\rho_0$ permutation finding (exploratory; cv_ratio 0.30–0.45).

ACKNOWLEDGEMENTS

This work uses the SPARC database (Lelli et al. 2016). Analyses were conducted using scikit-learn (Pedregosa et al. 2011), NumPy, pandas, SHAP (Lundberg & Lee 2017), SciPy, and the TAIJITU pipeline. Analyses were developed and verified using the Claude AI assistant (Anthropic, 2026). The parallel pipeline manuscript referenced as Melegh (2026 parallel) is a draft companion paper by the same author using complementary methodology. All numerical results are directly traceable to archived CSV files in the `taijitu_v14-v18_outputs.zip` series.

REFERENCES

1. Di Cintio, A., Brook, C. B., Macciò, A. V., et al. 2014, MNRAS, 437, 415
2. Haubner, K., et al. 2024, A&A [BIG-SPARC catalogue, in preparation at time of analysis]
3. Khelashvili, M., et al. 2022, MNRAS, 512, 2356
4. Lelli, F., McGaugh, S. S., & Schombert, J. M. 2016, AJ, 152, 157
5. Li, P., Lelli, F., McGaugh, S., & Schombert, J. 2020, ApJS, 247, 31
6. Lundberg, S. M. & Lee, S.-I. 2017, NeurIPS, 30
7. Melegh, J. G. 2026 (parallel). "Inner Zone Kinematic Fit Quality as the Primary Predictor of Rotation Curve Baseline Failure." Manuscript in preparation (same author, complementary methodology).
8. Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
9. Pedregosa, F., et al. 2011, JMLR, 12, 2825

APPENDIX A: NOTEBOOK VERSION HISTORY

Table A1. Notebook version history. Input panel: exact_explanatory_panel_v12.csv (175 × 157) — authoritative from v12 onward.

Version	Key additions	Primary finding / resolution
v1–v4	Baseline, calibration, mark+topology baseline	BDI-based separation established
v5–v8	Network community structure, role-cell ontology	Community structure adds signal
v9–v11	Exact backfill: halo, radial, sweep layers	Three backfill layers validated
v12	First exact explanatory builder; full panel 171×157	First exact gain: TRAIN→DEV AUC 1.000 (radial, full); 1.000 (radial, high-BDI) — single split only
v13 (N_REPEATS=2)	Feature stability cv_ratio; halo/radial shortlist; cross-project priors	full_non_HO supported (boot_lo = 0.967); high_BDI exploratory (boot_lo = 0.890)
v14 (N_REPEATS=8)	40-fold CV; CV-fold permutation importance; has_halo_backfill audit; PF DEV attempt	high_BDI supported ★ (boot_lo = 0.920); artefact identified; normalization_scale resolved
v15	Artefact exclusion; DC14 γ partial corr. (3 seeds); SHAP; BDI holdout regression; Spearman corr. matrix	10/10 MDI–SHAP concordance; DC14 γ n.s. own pipeline; holdout BDI ρ = 0.822; 6 halo / 223 radial high-corr. pairs
v16	Correlation pruning; holdout binary evaluation; GBM 100-split; enriched baseline issue identified	Holdout binary ρ = 0.836 ★; GBM radial +1.98% vs. MT; Burkert perm 0.145 (full)
v17	Pure obs. baseline (24 feat.); sign test vs. pure obs.; pruned block CV	Sign test 37/40 halo, 36/40 radial ($p < 10^{-6}$) ★; radial_pruned AUC = 0.984 ★ (series best); high-BDI gain +18.7 pp
v18	Baron–Kenny mediation (4 inner RMSE models); sensitivity analysis (5 confounders); perm re-run	BDI partial mediator ≈ 18% (exploratory); Hubble T confound identified; Burkert perm 0.301 (high-BDI)

APPENDIX B: DATA PROVENANCE

All numerical results reported in this paper are directly traceable to the following CSV files in the archived output ZIP series:

Table(s)	CSV file	ZIP archive
Tables 3, 6	baseline_repeated_cv_summary.csv	v14, v15, v16, v17, v18
Table 3 (pruned)	pruned_blocks_cv.csv	v17
Tables 4a, 4b, 13a, 13b	feature_stability_table_v14.csv	v14, v18
Table 5	permutation_importance_stability.csv	v14, v17, v18
Table 5b	corr_matrix_high_pairs_halo_exact_only.csv	v15
Table 5b (radial pairs)	correlation_pruning_report.csv	v16
Table 6	pure_obs_cv_comparison.csv	v17, v18
Table 7	gbm_100split_summary.csv	v16
Table 8	partial_corr_dc14_gamma.csv	v15
Tables 9a, 9b	shap_summary.csv	v15
Table 10 (binary)	holdout_binary_evaluation.csv	v16, v18
Table 10 (BDI reg.)	holdout_bdi_regression.csv	v15, v18
Tables 11, 12	mediation_baron_kenny.csv, mediation_sensitivity.csv	v18
All versions	manifest.json	v14–v18

Manuscript date: 2026-04-24. Panel: exact_explanatory_panel_v12.csv (175×157). Notebook lineage: v1–v18. The 60-galaxy HOLDOUT set was not used for binary classification fitting at any stage; the binary target `baseline_failure_10` was not computed for holdout galaxies by design. All p-values are two-sided unless otherwise noted. Bootstrap CIs are 95% percentile-method from 2,000 resamples of 40 CV fold-level AUC scores. The parallel pipeline (Melegh 2026 companion) is a separate manuscript by the same author using independently implemented analysis code on the same panel. *Not peer-reviewed.*